



**SENTIMEN ANALISIS TWITTER TERHADAP
PENYELENGGARAAN SEA GAMES 2021 HANOI
MENGUNAKAN ALGORITMA MACHINE LEARNING**

TUGAS AKHIR

Raihan Adam Pratama
41518010008

UNIVERSITAS
MERCU BUANA
**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022**



**SENTIMEN ANALISIS TWITTER TERHADAP
PENYELENGGARAAN SEA GAMES 2021 HANOI
MENGUNAKAN ALGORITMA MACHINE LEARNING**

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Raihan Adam Pratama

41518010008

UNIVERSITAS
MERCU BUANA
PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022

LEMBAR PERNYATAAN ORISINALITAS

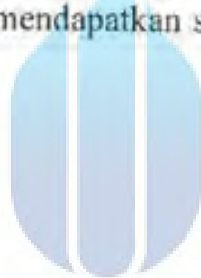
Yang bertanda tangan dibawah ini:

NIM : 41518010008

Nama : Raihan Adam Pratama

Judul Tugas Akhir : SENTIMEN ANALISIS TWITTER TERHADAP
PENYELENGGARAAN SEA GAMES 2021 HANOI
MENGUNAKAN ALGORITMA MACHINE LEARNING

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan di dalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



UNIVERSITAS
MERCU BUANA

Jakarta, 04 Juli 2022



Raihan Adam Pratama

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Raihan Adam Pratama
NIM : 41518010008
Judul Tugas Akhir : SENTIMEN ANALISIS TWITTER TERHADAP
PENYELENGGARAAN SEA GAMES 2021
HANOI MENGGUNAKAN ALGORITMA
MACHINE LEARNING

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 4 Juli 2022



Raihan Adam Pratama

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini

Nama Mahasiswa : Raihan Adam Pratama
NIM : 41518010008
Judul Tugas Akhir : SENTIMEN ANALISIS TWITTER TERHADAP
PENYELENGGARAAN SEA GAMES 2021
HANOI MENGGUNAKAN ALGORITMA
MACHINE LEARNING

Menyatakan bahwa :

1. Luanan Tugas Akhir saya adalah sebagai berikut :

No	Luanan	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan ✓
		Jurnal Nasional Terakreditasi	
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal	: CSRID (COMPUTER SCIENCE RESEARCH AND ITS DEVELOPMENT JOURNAL)	
	ISSN	: 2460-870X	
	Link Jurnal	: http://csrid.potensiutama.ac.id/ojs/index.php/CSRID/about/submissions#authorFees	
	Link File Jurnal Jika Sudah di Publish		

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui
Dosen Pembimbing TA



Sabar Rudiarto, S.Kom, M.Kom

Jakarta, 04 Juli 2022




Raihan Adam Pratama

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010008
Nama : Raihan Adam Pratama
Judul Tugas Akhir : Sentimen Analisis Twitter Terhadap Penyelenggaraan SEA Games 2021 HANOI Menggunakan Algoritma Machine Learning

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2022



(Achmad Kodar, Drs. MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010008
Nama : Raihan Adam Pratama
Judul Tugas Akhir : Sentimen Analisis Twitter Terhadap
Penyelenggaraan SEA Games 2021 HANOI
Menggunakan Algoritma Machine Learning

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2022



(Harni Kusniyati, M.Kom)

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010008
Nama : Raihan Adam Pratama
Judul Tugas Akhir : Sentimen Analisis Twitter Terhadap Penyelenggaraan SEA Games 2021 HANOI Menggunakan Algoritma Machine Learning

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2022



(Puji Rahayu, Dr, MT)

UNIVERSITAS
MERCU BUANA

LEMBAR PENGESAHAN

NIM : 41518010008
Nama : Raihan Adam Pratama
Judul Tugas Akhir : Sentimen Analisis Twitter Terhadap Penyelenggaraan SEA Games 2021 HANOI Menggunakan Algoritma Machine Learning

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 10 Agustus 2022

Menyetujui,



(Sabar Rudiarto, S.Kom., M.Kom)
Dosen Pembimbing

Mengetahui,



(Wawan Gunawan, S.Kom, MT)
Koord. Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Ka. Prodi Teknik Informatika

KATA PENGANTAR

Puji syukur kita panjatkan kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis bisa menyelesaikan tugas akhir yang berjudul “ SENTIMEN ANALISIS TWITTER TERHADAP PENYELENGGARAAN SEA GAMES 2021 HANOI MENGGUNAKAN ALGORITMA MACHINE LEARNING” dengan baik dan tepat pada waktunya.

Penulis menyadari bahwa tanpa bantuan dan bimbingan tidak dapat menyelesaikan tugas akhir ini dengan baik . Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua yang selalu memberikan support sehingga bisa menyelesaikan tugas akhir ini
2. Bapak Sabar Rudiarto,S.Kom., M.Kom. selaku Dosen Pembimbing Tugas Akhir saya yang telah membantu dan memberikan arahan dalam menyelesaikan tugas akhir ini
3. Bapak Emil R. Kaburuan Ph.D selaku Kepala Program Studi Informatika dan Dosen Pembimbing Akademik yang telah memberikan arahan, saran dan bimbingannya dalam menyelesaikan tugas akhir ini.
4. Bapak Wawan Gunawan S.Kom.,MT selaku Koordinator Tugas Akhir Prodi Informatika yang telah memberikan arahan selama periode tugas akhir
5. Teman-teman Tehnik Informatika 2018 yang selalu memberikan saran,kritik dan support sehingga dapat menyelesaikan tugas akhir ini tepat waktu
6. Sahabat Tonite yang selalu memberikan support saran dan semangat yang mendukung untuk menyelesaikan tugas akhir ini.

Akhir kata, penulis menyadari keterbatasan dan kemampuan dalam menyusun tugas akhir ini sehingga penulis mengharapkan masukan bersifat membangun sehingga tugas akhir ini lebih berkualitas dan bermanfaat bagi pembaca dan tentunya juga bagi penulis sendiri.

Jakarta,27 Juni 2022
Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR.....	iii
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI	v
LEMBAR PENGESAHAN	viii
ABSTRAK	ix
ABSTRACT.....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xii
NASKAH JURNAL	1
KERTAS KERJA.....	11
BAB 1. LITERATUR REVIEW.....	12
BAB 2. SOURCE CODE.....	31
BAB 3. DATASET.....	35
BAB 4. TAHAPAN EKSPERIMEN.....	40
BAB 5. HASIL SEMUA EKSPERIMEN.....	44
DAFTAR PUSTAKA	47
LAMPIRAN DOKUMEN HAKI.....	50
LAMPIRAN KORESPONDENSI	52

NASKAH JURNAL

■ 1

SENTIMEN ANALISIS TWITTER TERHADAP SEA GAMES 2021 HANOI MENGGUNAKAN ALGORITMA MACHINE LEARNING

Raihan Adam Pratama¹, Sabar Rudianto²

^{1,2}Universitas Mercu Buana, 021-584 0816,

³Jurusan Teknik Informatika, Fasilkom Mercu buana, Jakarta

e-mail: *¹ 41518010008@student.mercubuana.ac.id, ² sabar.rudiarto@mercubuana.ac.id

Abstrak

Ajang olahraga multievent di kawasan asia tenggara ialah SEA Games yang diadakan dua tahun sekali. Dalam pelaksanaan SEA Games, 11 negara di kawasan asia tenggara ikut berpartisipasi dalam event besar ini. Indonesia telah menjadi peserta SEA Games sejak 1977. Saat ini telah di selenggarakan SEA Games 2021 di Hanoi, Vietnam yang dilaksanakan pada tanggal 12 Mei 2022 hingga 23 Mei 2022. SEA Games ini seharusnya dilaksanakan pada November 2021, Namun harus di undur karena pandemic Covid-19. Dalam penyelenggaraan SEA Games ini masyarakat di Indonesia banyak memanfaatkan media twitter untuk menilai dan beropini tentang penyelenggaraan SEA Games 2021 Hanoi. Terjadinya opini ini didalam masyarakat, maka di perlukannya analisis sentimen. Analisis sentimen diperoleh di berbagai platform seperti twitter, instagram, facebook dan media sosial lainnya yang memiliki tujuan mengetahui sentimen apa yang terdapat di masyarakat terhadap penyelenggaraan SEA Games Hanoi 2021 melalui tweet pada media sosial twitter. Proses analisis dalam penelitian ini dilakukannya analisis sentimen mengenai penyelenggaraan SEA Games Hanoi 2021 yang dibagi menjadi tiga yaitu positif, negatif dan netral. Analisis sentimen pada penelitian ini menggunakan metode algoritma machine learning yaitu naïve bayes dan Support vector machine. Hasil uji akurasi naïve bayes mendapatkan akurasi 80% dan support vector machine mendapatkan akurasi 84%

Kata kunci— Analisis sentiment, SEA Games 2021, Twitter, Naïve, Bayes, Support Vector Machine

Abstract

The multi-event sports event in the Southeast Asian region is the SEA Games which is held every two years. In its implementation, the SEA Games are participated by 11 countries in the Southeast Asian region. Currently, the 2021 Sea Games will be held in Hanoi, Vietnam, which will be held from May 12, 2022 to May 23, 2022. The SEA Games were supposed to be held in November 2021, but had to be postponed due to the Covid-19 pandemic. In organizing the SEA Games, many people in Indonesia use Twitter to assess and give their opinion about the 2021 Hanoi SEA Games. The emergence of this opinion in society requires sentiment analysis. Sentiment analysis is obtained on various social media which has the aim of knowing what sentiments are in the community towards the SEA Games Hanoi 2021 through tweets on the media social. The analysis process in this research is sentiment analysis regarding the implementation of the 2021 SEA Games Hanoi which is classified into positive, negative and neutral. Sentiment analysis in this study uses machine learning algorithm methods naïve Bayes and Support vector machines. Naive Bayes accuracy test results get 80% and support vector machine get 84% accuracy.

Universitas Mercu Buana

Keywords— Analisis sentiment,SEA Games 2021,Twitter,Naive,Bayes,Support Vector Machine

1. PENDAHULUAN

Perkembangan teknologi informasi membuat manusia mengubah cara komunikasi terutama pada media sosial. Penggunaan media sosial di masyarakat untuk mengekspresikan opini atau pendapatnya pada media sosial seperti twitter,facebook,Instagram dan lain-lain. Hal mengekspresikan tersebut disebut dengan sentimen. Analisis sentimen merupakan sentimen,opini dan emosi yang tuangkan dalam bentuk tekstual.Opini dapat diartikan sebagai pendapat dalam menjelaskan suatu kecenderungan terhadap padangan atau gagasan seseorang yang masih bersifat tidak objektif [1].Opini bisa berupa mention pada media sosial ataupun artikel yang dimuat dalam situs atau blog.

Banyak penelitian terdahulu yang membahas analisis sentimen.Seperti penelitian yang membahas jumlah review komentar positif dan negative pada grub shopee pada twitter menggunakan algoritma naive bayes. Hasil yang di tunjukan pada komentar positif dan negative adalah seimbang. Data yang digunakan adalah 300 data dengan akurasi 97% [2].Penelitian lainnya tentang Pemilihan Gubernur DKI Jakarta 2017 yang sedang banyak diperbincangkan khususnya di media sosial Twitter. Preprocessing data seperti tokenisasi, cleansing dan filtering dan Algoritma yang digunakan adalah naive bayes dan SVM. Hasil yang di tunjukan dari analisis sentimen terhadap calon gubernur DKI Jakarta 2017 menggunakan naive bayes mendapatkan akurasi 95%,presisi 95%,recall 95%,TP Rate 96,8%,dan TN rate 84,6% [3]. Penelitian lainnya dengan algoritma naive bayes dan SVM untuk mengklasifikasikan opini publik dengan teks berbahasa Indonesia di twitter pada calon gubernur jawa barat periode 2018-2023.Hasil akurasi naive bayes mendapatkan 94% dan untuk SVM mendapatkan akuarasi 75,50% [4]. Penelitian lainnya mengenai analisis sentimen twitter BMKG nasional menggunakan naive bayes mendapatkan hasil akurasi 68,97% [5].Penelitian lainnya Analisis sentimen twitter kepuasan pengguna penyedia layanan seluler telekomunikasi dengan algoritma naive bayes dan lexicon based feature mendapatkan hasil akurasi 79%, presisi 65%,recall 97% dan f-measure 78% [6].

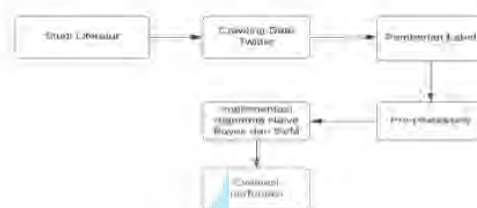
Ajang olahraga multievent di asia tenggara yaitu SEA games, Diadakan dua tahun sekali dan 11 negara di kawasan asia tenggara ikut berpartisipasi dalam penyelenggaraan SEA Games, SEA Games ini di naungi oleh Federasi Pesta Olahraga Asia Tenggara (SEAGF). Indonesia telah menjadi peserta SEA Games sejak 1977.Saat ini telah di selenggarakannya Sea Games 2021 di Hanoi,Vietnam yang dilaksanakan pada tanggal 12 Mei 2022 hingga 23 Mei 2022.SEA Games ini seharusnya dilaksanakan pada November 2021.Namun harus di undur karena pandemic Covid-19 [7].

Dalam penyelenggaraan SEA Games ini masyarakat memanfaatkan media twitter untuk menilai dan beropini tentang SEA Games 2021 Hanoi menurut pandangan masing-masing yang di luapkan dalam emosi. Emosi tweet pengguna twitter inilah disebut analisis sentimen. Analisis sentimen merupakan analisa mengenai pendapat atau opini terhadap masalah atau suatu objek yang diklasifikasikan menjadi positif,negatif dan netral.Dalam Permasalahan penelitian ini adalah masih belum diketahui sentimen tweet pada SEA Games 2021 Hanoi dan juga masyarakat terlihat sangat antusias sekali menyampaikan opini atau pendapat di media sosial twitter terkait SEA Games 2021 hanoi.

Dari permasalahan yang telah di jelaskan ,Maka peneliti tertarik untuk meneliti sentimen analisis terhadap SEA Games 2021 pada media sosial twitter.Penelitian ini menggunakan algoritma Naive bayes dan Support Vector Machine (SVM) dengan tools google colab untuk mendapatkan data di twitter dan menghitung tingkat akurasi atau perfoma algoritma dari data tersebut.

2. METODE PENELITIAN

Data penelitian ini diambil dari twitter dengan crawling data dengan kata kunci SEA Games 2022. Data yang diambil yaitu berupa opini-opini terkait dalam SEA Games 2021 Hanoi. Opini terbagi menjadi positif, negative dan netral. Metode yang digunakan dengan algoritma Naïve Bayes dan SVM. Pengambilan datanya menggunakan Google collab lalu melakukan labelling secara manual yang dimana data-nya telah terbagi menjadi tiga data yaitu opini positif, negative dan netral. Tahapan penelitian yang akan dilakukan di jelaskan pada gambar 1:



Gambar 1. Tahapan Penelitian

2.1 Studi Literatur

Langkah yang pertama yang dilakukan penulis adalah studi literatur. Studi literatur diperlukan untuk mencari suatu data, informasi atau referensi terkait dengan penelitian yang dilakukan. Sumbernya bisa berupa buku dan jurnal yang terkait.

2.2 Crawling Data

Crawling data merupakan proses pengumpulan data oleh sistem. Crawling data ini menggunakan API pada twitter. Langkah awal dalam crawling data membuat akun twitter dan selanjutnya melakukan pendaftaran pada <https://dev.twitter.com/apps/new>. Setelah registrasi dan mendapatkan kode akses API twitter, selanjutnya menggunakan tools Google colab untuk melakukan crawling data twitter.

2.3 Labelling

Labelling adalah proses penentuan kelas dalam tweet yang dilakukan secara manual. Label yang diberikan adalah positif negatif dan netral dalam penentuan sentimen memperhatikan ketentuan dalam setiap tweet apakah sentimen terkandung termasuk positif, negatif atau netral.

2.4 Pre-processing

Tahapan ini sangat diperlukan untuk mempersiapkan data agar dataset yang digunakan bersih dan siap untuk digunakan. Dalam preprocessing ada beberapa tahapan yang perlu dilakukan:

2.4.1 Cleaning

Tahapan cleaning adalah pembersihan setiap tweet untuk sesuatu yang tidak dibutuhkan dalam penelitian. Proses cleaning ini menghapus emoticon, username, dan URL.

2.4.2 Case Folding

Tahapan case folding merupakan proses menyesuaikan bentuk huruf besar atau menjadi huruf kecil

2.4.3 Word Stemming dan Lemmenting

Pada tahap ini penulis melakukan stemming dan lemmenting. Stemming merupakan Proses menghilangkan bentuk awalan atau akhiran kata,kata sambung,kata depan,kata ganti menjadi ke bentuk kata dasar yang sesuai. Lalu lemmenting merupakan proses untuk mengubah kata menjadi bentuk dasarnya.

2.4.5. Stopwords Removal

Tahapan ini merupakan proses penyaringan kata pada analisis sentimen yang tidak ada hubungannya dalam penelitian ini.

2.4.6. Tokenisasi

Tahapan ini adalah proses memecahkan kata menggunakan spasi menjadi token-token

2.5 Implementasi Algoritma

Tahapan ini adalah pengimplementasian ke metode algoritma yang akan digunakan.Setelah semua data telah di proses dan mendapatkan data bersih maka selanjutnya data akan di implementasikan menggunakan Algoritma yang diajukan.Berikut implementasi algoritma yang akan digunakan.

2.5.1 Naïve Bayes

Algoritma pembelajaran yang digunakan menangani permasalahan klasifikasi teks naïve bayes, suatu metode machine learning yang menggunakan probabilitas atau peluang.Perhitungan probabilitas mengambil pendekatan Bayesian.Pendekatan dalam algoritma ini mengkombinasikan probabilitas sebelumnya serta probabilitas kondisional dalam formula yang digunakan menghitung probabilitas tiap klasifikasi [8].

$$P(X|H) = \frac{P(X,H)}{P(H)}$$

MERCU BUANA

Keterangan :

X	: Data yang belum diketahui
H	: Hipotesis dari data class
P(H X)	: Probabilitas hipotesis nilai berdasarkan kondisi nilai X
P(H)	: Probabilitas hipotesis nilai H
P(X H)	: Probabilitas nilai X berdasarkan hipotesis nilai H
P(X)	: Probabilitas nilai X

2.5.2 Support Vector Machine (SVM)

Suatu metode yang digunakan dalam menganalisa data pada suatu pola pengklasifikasian. Metode ini dilakukan dengan memisahkan data optimal dengan membangun hyperplane dalam bentuk ruang input multi dimensi. Hyper-planes ini menjadi terbaik ketika di definisikan dalam mengukur batas-batas keanggotaan. Awalnya SVM digunakan dalam klasifikasi data dalam bentuk angka, akan tetapi ditemukan bahwa SVM sangat cepat dan baik dalam mengatasi masalah data teks. Data teks cocok pada algoritma ini dikarenakan sifat teks yang mempunyai dimensi yang tinggi dan memiliki fitur-fitur yang tidak relavan.

2.6 Evaluasi

Langkah terakhir melakukan evaluasi terhadap performa metode yang digunakan, Pada evaluasi ini peneliti bisa melihat apakah penelitian dilakukan mempunyai performa hasil yang baik atau tidak.

3. HASIL DAN PEMBAHASAN

Pada halaman ini hasil dan pembahasan mengenai penelian yang dilakukan. Pada penelitian ini juga menggunakan bantuan tools google colab untuk mengimplementasikan algoritma yang akan digunakan. Berikut hasil dan pembahasan penelitian ini menjadi beberapa tahapan sesuai dengan metodologi sebagai berikut:

3.1 Hasil Crawling data dan label

Proses crawling data telah berhasil dilakukan. Data yang diambil merupakan tweet dengan keyword "SEA Games 2022" dan total jumlah data terkumpul sebanyak 1497. Setelah data berhasil terkumpul maka selanjutnya dilakukan proses pemberian label. Pemberian label ini dilakukan secara manual. Berikut hasil crawling data dan labelling pada gambar 2

	waktu	username	teks	LABEL
0	20/05/2022 12:47	opsiin	hasil futsal sea games 2022 vietnam thailand	netral
1	20/05/2022 12:46	arehanawis	hasil futsal sea games 2022 vietnam thailand	netral
2	20/05/2022 12:33	arehanawis	hasil sea games christo alcala bon indonesia	netral
3	20/05/2022 12:39	opsiin	hasil sea games christo alcala bon indonesia	netral
4	20/05/2022 12:36	opsiin	petakul caitli caitalo modica futsal di futsal	netral
...
1492	18/05/2022 18:48	reza lawyer office	adwail adwail gangguan yoi sea games 2022	netral
1493	18/05/2022 19:33	ig wisapadaning	sea games timnas futsal indonesia amankah media	netral
1494	18/05/2022 19:11	Ribun poliklinik	update penitahan medal Indonesia di sea games	netral
1495	18/05/2022 18:47	Kumpang raiji	ah! leon ehawati sumbang erosa sea games futsal	positif
1496	18/05/2022 18:47	leontombak	lin kavitio indonesia sumbang erosa sea games futsal	positif

Gambar 2. Hasil Crawling data dan labelling

3.2 Hasil Pre-processing

Pre-processing merupakan mempersiapkan data agar data tersebut dapat digunakan .terdapat beberapa proses seperti case folding,word stemming dan lemmenting,stopword removal dan tokenisasi.

3.2.1 Case folding

Berikut hasil dari case folding pada gambar 3.

```

teks
hasil futsal sea games sikat vietnam thailand ...
hasil futsal sea games sikat vietnam thailand ...
hasil sea games christo aldila beri indonesia ...
hasil sea games christo aldila beri indonesia ...
pecatur cantik chelsie monica takluk di final ...

```

Gambar 3. Hasil case Folding

3.2.2 Tokenisasi

Hasil Tokenisasi pada penelitian ini pada gambar 4.

```

[ ] import nltk
nltk.download('punkt')
#Tokenization : In this each entry in the corpus will be broken into set of words
Corpus['teks'] = [word_tokenize(entry) for entry in Corpus['teks']]

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.

[ ] Corpus['teks']
0      [hasil, futsal, sea, games, sikat, vietnam, th...
1      [hasil, futsal, sea, games, sikat, vietnam, th...
2      [hasil, sea, games, christo, aldila, beri, ind...
3      [hasil, sea, games, christo, aldila, beri, ind...
4      [pecatur, cantik, chelsie, monica, takluk, di...
      ...
1492   [jadwal, siaran, langsung, voli, sea, games, t...
1493   [sea, games, timnas, futsal, indonesia, amanka...
1494   [update, perolehan, medali, indonesia, di, sea...
1495   [leki, febrri, ekawati, sumbang, emas, sea, game...
1496   [tim, karate, indonesia, sumbang, dua, emas, d...
Name: teks, length: 1497, dtype: object

```

Gambar 4. Hasil tokenisasi

3.2.3 Word stemming, lemmeting, dan stopwords removal

Berikut hasil dari word stemming, lemmeting, dan stopwords removal yang telah dilakukan pada gambar 5,6,7.

```

[ ] nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')
nltk.download('stopwords')
nltk.download('omw-1.4')
#Remove stop words, Non-Numeric and perform word Stemming/lemmating
#We need to tag words with noun, verb or adjective to pass to WordNetLemmatizer
tag_map = defaultdict(lambda : wn.NOUN)
# nltk.download('wordnet')
tag_map['J'] = wn.ADJ
tag_map['V'] = wn.VERB
tag_map['R'] = wn.ADV

```

Gambar 5. Proses stemming, lemmeting, dan stopwords removal

```

for index,entry in enumerate(corpus['teks']):
    # Declaring empty list to store the words that follow the rules for this step
    final_words = []
    # Initializing WordNetLemmatizer()
    word_lemmatized = WordNetLemmatizer()
    # pos_tag function below will provide the 'tag' i.e. (Noun(n) or Verb(v) or something else)
    for word, tag in pos_tag(entry):
        # below condition is to check for Stop words and consider only alphabets
        if word not in stopwords.words('indonesian') and word.isalpha():
            word_final = word_lemmatized.lemmatize(word,tag_map[tag[0]])
            final_words.append(word_final)
    # The final processed set of words for each iteration will be stored in 'text_final'
    corpus.loc[index,'text_final'] = str(final_words)

```

Gambar 6. Proses stemming, lemmeting, dan stopwords removal

```

[ ] #lets see the data
Corpus['text_final']
0      ['hasil', 'futsal', 'sea', 'game', 'sikat', 'v...
1      ['hasil', 'futsal', 'sea', 'game', 'sikat', 'v...
2      ['hasil', 'sea', 'game', 'christo', 'aldila', ...
3      ['hasil', 'sea', 'game', 'christo', 'aldila', ...
4      ['pecatur', 'cantik', 'chelsie', 'monica', 'ta...
...
1492   ['jadwal', 'siaran', 'langsung', 'voli', 'sea'...
1493   ['sea', 'game', 'timnas', 'futsal', 'indonesia...
1494   ['update', 'perolehan', 'medali', 'indonesia'...
1495   ['eki', 'febri', 'ekawati', 'sumbang', 'emas'...
1496   ['tim', 'karate', 'indonesia', 'sumbang', 'ema...
Name: text_final, Length: 1497, dtype: object

```

Gambar 7. stemming, lemmeting, dan stopwords removal

3.3 Hasil implementasi Algoritma dan evaluasi

Setelah data selesai di pre-processing dan data siap digunakan. Maka selanjutnya masuk ketahap implementasi algoritma dan evaluasi. Metode ini menggunakan algoritma machine learning yaitu Naive bayes dan SVM. Hasil implementasi algoritma dan evaluasi ini dibagi menjadi dua bagian Confusion Matrix dan Classification Report. Hasil tersebut penulis bagi menjadi dua sebagai berikut:

3.3.1 Confusion Matrix

Pada bagian ini merupakan evaluasi menggunakan Confusion Matrix dan pada percobaan menggunakan algoritma Naive bayes dan SVM. Berikut confusion Matrik di jelaskan pada table 1.

Tabel 1. Confusion Matrix Naive bayes dan SVM

		Predicted			
			Positif	Negatif	Netral
Actual Table	Naive Bayes	Positif	1	0	41
		Negatif	0	0	32
		Netral	0	0	301
	SVM	Positif	11	0	31
		Negatif	0	2	30
		Netral	1	0	300

3.3.2 Classification Report

Pada bagian ini merupakan hasil evaluasi menggunakan classification report menggunakan algoritma Naive bayes dan SVM. Berikut classification report di jelaskan pada tabel 2 dan 3.

Tabel 2. Classification Report Naïve Bayes

Naïve Bayes				
	Precision	Recall	F1-Score	Support
Negatif	0.00	0.00	0.00	32
Netral	0.80	1.00	0.89	301
Positif	1.00	0.02	0.05	42
Accuracy	0.81			375
Macro avg	0.60	0.34	0.31	375
weighted avg	0.76	0.81	0.72	375

Tabel 3. Classification Report SVM

Support Vector Machine				
	Precision	Recall	F1-Score	Support
Negatif	1.00	0.06	0.12	32
Netral	0.83	1.00	0.91	301
Positif	0.92	0.26	0.41	42
Accuracy	0.83			375
Macro avg	0.92	0.44	0.48	375
weighted avg	0.86	0.83	0.78	375

4. KESIMPULAN

Dari penelitian ini maka dapat disimpulkan kedua Algoritma Naïve Bayes dan Support Vector Machine memiliki hasil performa cukup bagus dalam menganalisis sentiment twitter pada SEA Games 2021 Hanoi. Hasil yang didapatkan akurasi dari naïve bayes 81%, sedangkan SVM mendapatkan akurasi sebesar 83%. Hal ini juga menunjukkan bahwa algoritma SVM lebih baik dari pada naïve bayes dalam hasil performa pada penelitian analisis sentimen tweet pada SEA Games 2021 hanoi ini.

5. SARAN

Saran saya untuk penelitian selanjutnya lebih baik menggunakan lebih banyak data dan dalam pre-processing juga dilakukan lebih baik lagi dan juga bisa mengganti algoritma dengan yang lain sehingga bisa mendapatkan hasil yang maksimal.

UCAPAN TERIMA KASIH

Penulis berterimakasih yang sebanyak-banyaknya kepada kedua orang tua memberi dukungan, motivasi dan bantuan sehingga penulis dapat menyelesaikan penelitian ini dengan lancar dan hasil yang baik

DAFTAR PUSTAKA

- [1] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.
- [2] A. H. Ruger, M. Suyanto, and M. P. Kurniawan, "Sentimen Analisis Pelanggan Shopee di Twitter dengan Algoritma Naive Bayes," *J. Inf. Technol.*, vol. 1, no. 2, pp. 26–29, 2021, doi: 10.46229/jifotech.v1i2.282.
- [3] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," 2017. [Online]. Available: <https://t.co/jrvaMsgBdH>.
- [4] C. Rafiqah, "Analisis Sentimen Pada Media Sosial Twitter Terhadap Tokoh Publik Peserta Pilpres 2019," *Jurnal Ilmu Komputer dan Teknologi Informasi*, Vol 12, No. 1, 2020, pp 79-86, doi: 10.18860/mat.v12i1.8356.
- [5] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [6] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1725–1732, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628>.
- [7] Erianto, D., 2022, SEA Games: Sejarah, Cabang, Peserta, dan Prestasi Indonesia <https://kompaspedia.kompas.id/baca/paparan-topik/sea-games-sejarah-cabang-peserta-dan-prestasi-indonesia>, diakses tgl 03 Juli 2022
- [8] T. T. Widowati and M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine," *Simetris J. Tek.*

Mesin, Elektro dan Ilmu Komput., vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.

- [9] Saruni Dwiasnati and Yudo Devianto, “Classification of forest fire areas using machine learning algorithm,” *World J. Adv. Eng. Technol. Sci.*, vol. 3, no. 1, pp. 008–015, 2021, doi: 10.30574/wjaets.2021.3.1.0048.
- [10] S. Dwiasnati and Y. Devianto, “Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir,” *Pros. SISFOTEK*, pp.202207,2021,[Online].Available:<http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/283>.



KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul "Sentimen Analisis Twitter Terhadap SEA Games 2021 Hanoi Menggunakan Algoritma Machine Learning". Kertas kerja berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat/atau disertakan di artikel jurnal. Di dalam kertas kerja ini disajikan: literature review, dataset yang digunakan, source code, dan hasil eksperimen secara keseluruhan.

